



A11103 811531

NISTIR 4892

NIST
PUBLICATIONS

RESEARCH CONSIDERATIONS REGARDING FBI-IAFIS TASKS & REQUIREMENTS

**R. McCabe
C. Wilson
D. Grubb**

U.S. DEPARTMENT OF COMMERCE
Technology Administration
National Institute of Standards
and Technology
Gaithersburg, MD 20899

QC
100
.U56
4892
1992
C.2

NIST

RESEARCH CONSIDERATIONS REGARDING FBI-IAFIS TASKS & REQUIREMENTS

R. McCabe
C. Wilson
D. Grubb

U.S. DEPARTMENT OF COMMERCE
Technology Administration
National Institute of Standards
and Technology
Gaithersburg, MD 20899

July 1992



U.S. DEPARTMENT OF COMMERCE
Barbara Hackman Franklin, Secretary

TECHNOLOGY ADMINISTRATION
Robert M. White, Under Secretary for Technology

NATIONAL INSTITUTE OF STANDARDS
AND TECHNOLOGY
John W. Lyons, Director

CONTENTS

1.	BACKGROUND	1
2.	INTRODUCTION	1
3.	SCOPE AND PURPOSE	2
4.	OVERVIEW OF THE C/M REPORT	2
5.	ANALYSIS OF THE C/M REPORT	4
6.	DISAGREEMENTS WITH THE C/M REPORT	5
6.1	CLASSIFICATION	5
6.2	VERIFICATION	5
6.3	COMPETITIVE ALGORITHMS	6
7.	C/M RECOMMENDED RESEARCH	6
7.1	TRAINABLE NEURAL NETWORKS	6
7.2	RELAXATION ALGORITHMS	6
7.3	SCALE-SPACE ALGORITHMS	7
7.4	EXAMINER EXPERTISE	7
7.5	GRAPH MAPPING	7
8.	CLASSIFICATION CONSIDERATIONS	8
8.1	CLASSIFICATION PROBLEM	8
8.2	SHORT-TERM SOLUTIONS TO CLASSIFICATION	10
8.3	LONG-TERM SOLUTIONS TO CLASSIFICATION	12
8.4	CLASSIFICATION RECOMMENDATION	13
9.	MINUTIAE DETECTION ISSUES	13
10.	RESEARCH DATABASES	13
11.	OTHER IAFIS TASKS	14
11.1	TEXTUAL DATA	14
11.2	IMAGE QUALITY	15
11.3	GRAYSCALE REQUIREMENTS	15
11.4	BACKUP REQUIREMENTS	15
11.5	OPTICAL STORAGE	16
11.6	LOCAL AND WIDE AREA COMMUNICATION STANDARDS	16
11.7	SOFTWARE STANDARDS	17

12. SUMMARY & CONCLUSIONS	17
REFERENCES	18

TABLES

1 MATCHERS REQUIRED	9
---------------------------	---

1. BACKGROUND

The FBI's Identification Division (ID) is responsible for maintaining a ten-print fingerprint record repository, and providing timely ten-print and latent identification services to requesting agencies. For years the FBI held a position of leadership in these identification services. In the late 1970's, computer databases were created containing subjects' names, associated physical and demographic data, and fingerprint information including characteristic or minutiae data. An Automated Fingerprint Identification System (AFIS) was installed and became operational. Automated searches were performed against these databases and identifications resulted.

During the 1980's, state and local governments began procuring their own AFISs. These were commercial, state-of-the-art systems designed with database sizes and processing capabilities to handle the workload demands at that time. Provisions were also included to accommodate reasonable growth. Users were very satisfied with the performance offered by these systems, which were not hampered by fast growth.

Meanwhile, demands on the AFIS and other automated databases in the ID grew at a rapid rate. The daily candidate submissions added to the databases and the number of daily search requests surpassed the intended system design capabilities. As a result of this high demand, turnaround times for each response increased to an unreasonable level. Manual operations for entering each subject's descriptive information and their fingerprint classification codes from the fingerprint card were major contributors to these increased delays. Turnaround times were further increased due to the necessity to physically transport the fingerprint card associated with each submission through the many processing stations involved in the identification operation.

As a result, users became increasingly dissatisfied with FBI provided services and identification response times. The former position of FBI leadership and success caused the user community to demand improvement of FBI services and faster turnaround times for the identification operation.

2. INTRODUCTION

In order to improve the level of service to its users, the FBI is engaged in a revitalization effort. One of the end products will be a state-of-the-art Integrated Automated Fingerprint Identification System (IAFIS). This system is to provide a smooth and timely flow of work from the name search operation through the fingerprint classification, scanning, matching, and verification processes. It is intended to provide identification services to National Crime Information Center (NCIC) subscribers. When completed, the IAFIS will support a paperless environment by processing electronic images of fingerprints and other electronic data stored in ASCII format. Manual handling of paper documents throughout the identification processes will be eliminated. Without increasing its staff size, the Identification Division (ID) will be able to handle an ever

increasing workload. While the current response time for a search request is weeks, IAFIS is being designed to provide a two hour response time for criminal inquiries.

This revitalization effort is a monumental task. The FBI has solicited input from various sources in order to provide its users with the best possible system. One such source is the Advisory Policy Board (APB), which is part of the NCIC. Suggestions offered to the FBI by this group of individuals has been largely based on their own successful solutions to providing identification services. The FBI also solicited an independent view of the current state of the revitalization effort. A study performed by Carol Crawford and Eric Mjolsness for the FBI provided such an independent view.¹

3. SCOPE AND PURPOSE

This paper is primarily intended to provide a review of the programmatic needs and the directions of technical efforts that the FBI's ID should consider. Potential pitfalls contained in the Crawford and Mjolsness (C/M) report will be identified and some alternative methods for solving these problems will be provided. Candidate areas of IAFIS research, including those suggested by the C/M report, will be examined. Although several suggestions and viewpoints will be introduced, this paper is not intended to be a system analysis of the developing IAFIS project.

Section 4 of this paper presents an overview of the three alternative IAFIS design approaches available to the FBI at this time and the recommendations described in the C/M report. Section 5 provides a general analysis of the report. Major points of disagreement with the C/M recommendations are presented in Section 6 followed by a closer look at their research recommendations and minor points of difference in Section 7. An in-depth examination of the classification problem and possible short-term and long-term solutions are presented in Section 8. The importance of reliable minutiae detection is presented in Section 9. Section 10 contains a discussion of the need for research databases. Chapter 11 contains a brief discussion of several other topics of concern. Finally, a summary of finding will be presented in Section 12.

4. OVERVIEW OF THE C/M REPORT

The C/M report states that the IAFIS project is a necessary one, but intrinsically risky. Various critical issues inherent in IAFIS support this statement. Committing to new technologies which utilize innovative methods in the various aspects of automatic fingerprint identification could greatly benefit the FBI's revitalization effort. However, it could also result in failure and disaster! Alternatively, the FBI could limit itself to the use of conventional AFIS techniques which have proven adequate for smaller systems. But this approach could also lead to disaster when the techniques utilized are scaled up to the size required by the FBI. What level of commitment should the FBI devote to pioneering new technologies for use in their IAFIS? At this time, the FBI is faced with three alternative viewpoints on the development of the IAFIS.

First, "it appears that competition from state run AFISs has driven the recommendations and resulting structure for the FBI IAFIS." This APB-supported approach suggests that by simply sizing the FBI's system linearly from a smaller system will enable the FBI's operation to run at least as well as the smaller system. If the workload and file size of a system with a successful AFIS operation is one-tenth that of the FBI, then ten times those resources will be needed to fulfill FBI requirements. This argument contains the assumption that the relationship of the required resources between the two organizations is a linear one. As will be described in the classification discussion, the square of the amount of resources is a closer approximation.

If the FBI totally commits to this course of using a conventional IAFIS, the risk of obsolescence is present, especially if the projected rate of growth has been underestimated. Furthermore, it may not be technically feasible, using a conventional IAFIS, to take advantage of future technological developments associated with fingerprint identification. Commercial systems developed and installed during the 80's will soon be ready for major upgrades. These enhanced systems will probably be based on future technological advances. If the FBI commits to a conventional AFIS technology now, they will once again find themselves in the position of technological obsolescence as new systems come into existence.

The second approach that can be taken is for the FBI to commit itself to take the lead and develop new and innovative methods in the field of fingerprint identification. However, this would be very expensive and inherently risky. If better approaches to identification could not be developed in the given time frame, then the FBI could find itself operating in the same environment as it currently does.

Both approaches have risks of failure. Consequences of such failure include possible loss of function, excessive actual costs, and lost opportunity costs. The likelihood of these consequences is not remote and the effects are undesirable at best.

A third alternative is outlined in the C/M report. Due to the time restraints already in place, the report recommends that the FBI proceed with a modified conventional IAFIS plan, but not totally commit to it without specific adjustments. As this modified plan is being implemented, research efforts should be in progress directed toward improving the operation of the tasks associated with the identification process.

Modifications to the current plan as outlined in the report cover three basic areas. First, task modularity must be enforced to ensure continual openness to new technological developments. The C/M report splits the identification operation into the four basic tasks of fingerprint classification, image processing (including minutiae detection), fingerprint matching, and verification of candidate identifications. This recommendation is in contrast to the APB, which recommended an integrated, single vendor system to perform all of the above mentioned tasks. Enforcing task modularity allows new developments to be integrated into the existing system. Accepting a single system approach would discourage any major improvements to the initial design of the operation.

Second, the C/M report recommends that the FBI provide outside researchers with access to their fingerprint databases and details of existing processing methods. This will encourage serious outside research into various aspects of fingerprint technology. In the past, researchers have not had enough real data to enable them to perform any meaningful tests of their work. The availability of adequately sized databases can result in unsolicited free research efforts.

Third, the approach to classification methods should remain an open issue. The method used to classify fingerprints shall be a critical factor for the efficient operation of the IAFIS. Improvements in the classification process could decrease the number of matcher units required, decrease the time required for determining candidate identifications, and generally benefit the overall performance of the system. Using an open systems approach, as better algorithms or methods of classifying fingerprints are developed, these algorithms can be integrated into the system. Knowledge gained and techniques developed from research accomplished in this area may potentially be transferred to other fingerprint identification tasks.

The report also recommends that the FBI initiate both medium range and long range programs of research. This research would address the four main tasks of classification, image processing, matching, and verification using emerging techniques such as trainable neural networks. This is a very sound approach. The report states the fact that during the past decade little serious research has been directed toward fingerprint identification. Meanwhile, huge strides have been made in mathematics, computer science, and related fields. Many opportunities exist for applying the advances made in these new technologies to fingerprint identification tasks. Consequently, a relatively small amount of applied research, aimed at transferring the advances made in these technologies to fingerprint identification tasks, may result in tremendous improvements for overall IAFIS operation. Applied research in the proper areas, such as classification, could improve accuracy and decrease the cost of the system.

5. ANALYSIS OF THE C/M REPORT

From the content and ideas put forth in this C/M report, the authors appear to have a firm grasp of the various considerations and ramifications involved with the development of this IAFIS project.

The majority of their conclusions and recommendations are well founded and should be supported by the FBI. Any system installed should use the open system approach. This will permit the FBI to integrate future automated fingerprint identification enhancements into IAFIS as they become available and are proven to be superior to existing methods. Short, medium, and long term research projects in support of the various identification tasks should be initiated and continually supported. Classification of fingerprints is a prime candidate for this type of research. It also follows that large databases should be made available to outside researchers.

If the open systems approach suggested by the C/M report is adopted, standard sets of images and data will be important for the comparison of different problems. For example, if

classification methods are to be evaluated, the same sets of images should be used by all researchers for developing, testing, and evaluating a new algorithm. This will require that the FBI or some other organization prepare these data sets and specify the formula or criteria to be used for evaluation. Each subsystem that becomes open will require such sets of data. The classification subsystem will require sets of image data, image processing may require different sets of images, minutiae matching will require minutiae sets, etc.

6. DISAGREEMENTS WITH THE C/M REPORT

Although we support the majority of C/M report, we do disagree with certain aspects of it. The following paragraphs discuss those issues with which we have major concern. Minor points of disagreement are addressed in the paragraphs which discuss specific areas of research suggested by the C/M report.

6.1 CLASSIFICATION

The APB, the Mitre report², and the C/M report all agree that the NCIC system of fingerprint classification should be abandoned at this time in favor of classifying fingerprints into one of seven categories. However, using this classification technique results in a bin structure which is too coarse. While state and local agencies have used this approach successfully, these agencies have database sizes and daily workloads which are only a small fraction of those found at the FBI.

Since the FBI will have several times the number of fingerprints in its files and will be making several times the number of searches per day against these files, there will be an exponential increase in the number of searches per day. This is a major problem that will likely require hundreds of matchers, unless finer classifications can be made so as to reduce the bin sizes.

Currently, there are no proven fully reliable automatic classifiers in operation. But additional semi-manual classifying techniques are feasible, which would require much less classifier proficiency than the current NCIC classification. These techniques could result in large improvements in the initial classification process. This suggestion is not meant as a replacement for additional classification research. It is simply intended as an interim measure to be used until a better classification scheme is developed and proven.

6.2 VERIFICATION

The report refers to automating the identification verification process. Currently this function is performed by fingerprint examiners. Final verification should always be performed by skilled, experienced, human experts. If more emphasis and research is directed toward areas of image processing and matching, the number of candidate drops sent to human verifiers will be reduced. Therefore, only the most probable candidate drops will be forwarded to human

experts and the need for automatic verification will be diminished.

6.3 COMPETITIVE ALGORITHMS

As part of the conventional IAFIS system, the C/M report suggests "an annual review to determine which fingerprint identification algorithms should actually be used in IAFIS during the coming year." Every algorithm used to perform any of the identification tasks would be a candidate for this competitive replacement. Although the ideas of openness and task modularity should be supported, conducting an annual algorithm competition seems like an over-kill approach. As improved methods for performing operations are identified and determined to be superior, serious consideration should be given to integrating them into the current system. However, annual competitions would consume great amounts of FBI human resources devoted to creating new sets of research databases, conducting the competitions, and evaluating the results. Potential improvements are unlikely to justify the expenditure of resources.

7. C/M RECOMMENDED RESEARCH

The report recommends that the FBI initiate and promote applied research in fields of mathematics and computer science that could benefit the major fingerprint identification tasks. Advances in each of the research efforts listed below have the potential of benefiting one or more of the classification, image processing, or matching tasks required in the fingerprint identification process.

7.1 TRAINABLE NEURAL NETWORKS

The C/M report implies a comparison between trainable neural networks used for optical character recognition (OCR) and the tasks associated with fingerprint identification tasks. One point that appeared to have been overlooked in this discussion was the fact that fingerprints are natural objects. They were not designed, as are even the worst examples of handwriting, to be easy to recognize. The similarity to the AT&T or other OCR approaches may be of limited use.

7.2 RELAXATION ALGORITHMS

Due to the elastic nature of skin, there is an intuitive appeal to perform additional research in areas of relaxation algorithms. These algorithms, coupled with mapping and other principles from the field of graph theory, provide the opportunity for potentially significant advances to be made in the image processing and matching tasks. Although the C/M report points out that these approaches are computation-intensive, the use of parallel processing should reduce this concern. Matchers at the FBI employing the basic M-82 algorithm already take advantage of this general approach.

7.3 SCALE-SPACE ALGORITHMS

This class of algorithms includes Bipyramids and Gabor filters and wavelets which may be applied to image processing, matching, and classification tasks. These techniques are recommended by the C/M report.

Although referenced in the report, the application of wavelets for use in image compression deserves more attention. One of the goals of the IAFIS is to have all fingerprint images on-line and available for examiner verification. Because of this need, an efficient, high compression, low-loss algorithm is a requirement of the system.

At the NIST workshop³, the consensus of the attendees was that the compression algorithm of choice would be wavelet-based. The Wavelet Scalar Quantization (WSQ) Open Architecture was adopted as the framework for grayscale fingerprint compression and decompression. It was further decided that the FBI develop a parameterization technique certified for fingerprint image quality to be made freely available to the criminal justice community. This approach is intended to accommodate future enhancements, which may improve the quality and compression features.

However, one consideration to bear in mind when dealing with these scale-space algorithms is that much of the information required by fingerprint examiners is contained in the high-frequency components of the image. If the spacial resolution suffers too much of a decrease, valuable information will be lost to the verification process.

7.4 EXAMINER EXPERTISE

The active use of examiner expertise in the design of expert systems may be an untapped resource. The FBI should pursue further research along these lines. Input from the fingerprint examiner community may be of significant assistance in developing a more efficient IAFIS. Understanding the fingerprint examiners' perception of image quality and modelling the methodologies used in the verification process can potentially impact development of the IAFIS in a positive manner. Benefits to image processing, classification, and matching functions may also be a result of proceduralizing this expertise.

7.5 GRAPH MAPPING

Some of the concepts described for the graph mapping research have already been initially explored. Ridge counts to nearest minutiae have been used by NEC and Logica with positive results. Previous work using minutiae clusters as part of the matching process has also been successful. However, unless false and missing minutiae are kept to a minimum, only limited success may result.

The C/M report suggests that this approach can be used to improve classification tasks. However, in current work at NIST on fingerprint recognition, 90% classification was achieved with no registration. This implies that the topological mapping methods suggested in the report

cannot change attained results of current classification techniques by more than 10 percent.

However, feature compression has proved to be important in the classification problem. The researchers at AT&T and others at the recent NIPS⁴ conference have shown that in character recognition systems, minimizing the dimension of the network is very important. Most existing minimization methods make extensive use of Euclidian norms for learning and minimization. This is in contrast to the C/M report which recommends that research involving Euclidian distances be avoided.

8. CLASSIFICATION CONSIDERATIONS

A fast and accurate scheme for classifying fingerprints is crucial for a successful IAFIS system. Any classification approach to be used must have the ability to reliably select a minimum number of ten-print records from the master file to be processed by the matcher system. Failure to accomplish this objective results in routine search inquiries being compared against an unnecessarily large portion of the master file.

The conventional IAFIS system planned for operation in Clarksburg, WV is intended to conduct ten-print searches using a fingerprint classification system that is no more detailed than pattern level. This requirement may result in a system that is much larger and more costly than necessary with no accompanying assurance of success.

8.1 CLASSIFICATION PROBLEM

An adequate number of classification categories must exist in order to narrow down the ten-print master file. Lack of a manageable list of possible candidates on which to perform a technical search will require the use of hundreds of matchers. According to the FBI's own projections, the size of the ten-print master file is projected to grow from 28 million records by the end of 1994, to 64 million records at the end of the year 2008. A classification scheme based solely on seven pattern types will necessitate searching a large portion of this master file for many of the daily workload inquiries.

To be effective and efficient, a classification scheme requires a relatively uniform distribution of pattern types. However, the problem with using only seven gross pattern types for classification is the natural occurrence of a nonuniform distribution of the fingerprint pattern types.

Based on classifying all ten fingers by pattern type, the classifications for 80% of the subjects in the FBI's master file can be spread over 895 unique bins. The most frequently occurring pattern types will form the largest classification bins. However, almost 20% of the FBI's master file of ten-print cards are contained within ten of these bins. This is not a desirable situation. Because the incoming workload reflects the most frequently occurring pattern types, this portion

of the file will have to be searched most often.

Distribution statistics available from the FBI, show that approximately 6.1% of all fingerprint records contain ten ulnar loops. It is expected that by the end of 1995 over 31.7 million criminal fingerprint cards will be on file. The all-ulnar loop classification bin for this population will exceed 1.93 million individuals. By the same period, it is anticipated that the average daily search request to the minutiae matchers will exceed 44 thousand. Of these, approximately 2700 searches will be expected to contain ten ulnar loops and will be compared to the all-ulnar bin of the master file.

In order to process this workload within a 12 hour workday, the system will have to process over 225 ulnar loop search requests per hour. Each request may involve the comparison of four finger images or 900 searches per hour against each of the 1.93 million individuals in the all-ulnar bin. Using matchers capable of performing 1000 finger comparisons per second, over 483 matcher storage device configurations may be required working in parallel to handle the workload associated with the all-ulnar bin.

Table 1 summarizes the matcher requirements for the all-ulnar bin (additional matchers would be needed to handle the other bins) for the years 1995, 2000, and 2008. In the calculations for the number of matchers required, the arrival rate of search requests was based on the average hourly arrival rate. No provision was factored into the calculations for a peak arrival rate that could be considerably different than the average. For example, if the peak arrival rate would be double the average, twice as many matcher units as listed in the table would be required in order to guarantee the 1.5 hour turnaround objective.

<u>YR</u>	<u>MASTER FILE SIZE</u>	<u>ULNAR BIN</u>	<u>DAILY REQ</u>	<u>ULNAR /DAY</u>	<u>REQ /HR</u>	<u>SRCH /HR</u>	<u>NUMBER MATCHERS</u>
1995	31,714,000	1,934,554	44,253	2,700	225	900	483
2000	42,977,000	2,621,597	45,540	2,778	232	928	676
2008	63,518,000	3,874,598	51,725	3,155	263	1052	1,132

TABLE 1 MATCHERS REQUIRED

Therefore, systems much smaller than that required by the FBI can function successfully using only a pattern level classification. However, the matcher workload increases as the product of the increase in file size and the increase in the number of searches. Normally the number of searches will increase in proportion to the file size. Therefore, the matcher workload will increase as the square of either the increase in file size or number of searches. Since the bin sizes for pattern level classification are not equal in size, a small percentage of the 895 bins will be strongly impacted by any increase in database size or daily search requests.

Consideration must also be given to the bandwidth capabilities required to move data from the storage devices to the matchers. Generally, the use of pattern level classification requires many more matches to be performed for a given search request. Minutiae data from all the candidate subjects must be transferred to the matchers for comparison. Assuming that minutiae-based matchers will be used, each minutia will require on the order of eight bytes of storage. If there are 100 minutiae in the average fingerprint record, a 1,000 fingerprint per second matcher will require a sustained average data transfer rate of 800 kilobytes per second from the storage device to the matcher.

As both the workload and the size of the master file increase substantially, faster matchers, more matchers, or both will be required to process daily workloads. The demand on bandwidth will increase proportionately to the speed of the matchers, the number of matchers, the number of daily search requests, and the size of the master file. Although the 800 kilobytes per second may be currently attainable, substantial increases above this may be required to meet performance specifications for throughput and response time objectives. The bandwidth available may be insufficient to meet projected requirements.

Due to assumptions that were made in the above discussion and the use of other discriminators, the actual number of matchers required in an operational system may be less than calculated. However, as an AFIS is sized up, the number of matchers required is an exponential rather than a linear function of file size or workload.

8.2 SHORT-TERM SOLUTIONS TO CLASSIFICATION

Using a fingerprint classification system based only on gross pattern type will be a costly approach. Any refinement of this classification system would result in major cost and efficiency benefits. For example, if ulnar loops were to be only subdivided into two categories such as high and low ridge count ulnar loops for each of the ten fingers, this would have the potential of breaking up the all ulnar loop classification bin into 1024 smaller bins. The bins would not be of uniform size and referencing would be required in many cases. Nevertheless, this modest improvement in classification detail would have a dramatic impact on bin size and on matcher workloads as well as on the total IAFIS system size and cost.

8.2.1 SEMI-AUTOMATIC RIDGE COUNTING

A bottleneck in the current fingerprint identification system is the coding of the NCIC class codes by the fingerprint classifiers. One of the more costly processes of this operation in terms of manpower is the counting of ridges in loops. The counting process could possibly be semi-automated if each image is presented so that a human could point out the starting and ending points for the ridge count. Ridges crossing an imaginary straight line between the starting and ending points could be automatically located and counted. A plan for such an approach could be developed and tested using a workstation equipped with a pointing device such as a trackball with buttons and a display terminal.

As each fingerprint image from a ten-print record is displayed, the classifier would click a button to indicate the gross pattern type. If the type "clicked" was a loop, the classifier would position one pair of cross-hairs at the core and another pair at the delta of the loop. Each position would be registered by use of a button. By analyzing the pixels between these two points, the ridge crossings could automatically be determined by a computer program. The effort required to develop a workstation to test the feasibility of this approach would be a fairly short term research project. If this research application proved successful, then it could be implemented on a test basis at the FBI using equipment that could be associated with the image output of the present scanners. The motivation for developing this approach is to provide an interim technique for classifying fingerprints while a fully automatic long-term solution is being developed. Although requiring more human judgement than pattern recognition alone, it would provide a better approach to classification by reducing the number of candidate ten-prints that must be sent to a matcher.

8.2.2 CORE-TO-DELTA DISTANCES

Rather than classify loops by ridge counts, they could also be classified based on the core-to-delta distances. The distances would be coded as either small or large. By use of this simple technique, 1024 bins would be established for the all ulnar classification. Of course there would be several references made to both the small and large categories. But one or two additional distance categories could also be included to increase the number of bins and decrease the bin sizes. A short term research effort to collect statistics on core-to-delta distances would have to be conducted. An approach similar to the one for counting ridges could be implemented to determine these statistics.

An experiment to determine the feasibility of this approach might be implemented by inscribing three concentric circles on clear plastic mounted on the examiners magnifying loupe. The circles should be about 2 mm, 8 mm, and 10 mm in diameter. The exact diameters would be determined from the statistics derived above. The inner, smallest of these circles would enclose the core of the fingerprint and the others used to measure the distance to the delta. If the delta was inside the next smaller (8 mm) circle, the print would be classified small. If it was outside the largest circle it would be classed large. Anywhere else it would be referenced to both the small and large classes. It is also possible to perform the same experiment by positioning these circles over images displayed on a monitor.

8.2.3 VOICE INPUT

Although not designed specifically as a short term project, the use of voice input to speed up the classification process should be investigated. By eliminating the task of writing or typing the classification code for each finger, the classifier is free to process more images than otherwise possible. It may be worthwhile for the FBI to consider sponsoring research to investigate voice input for classification purposes.

8.3 LONG-TERM SOLUTIONS TO CLASSIFICATION

Ultimately, the best long-term solution to this problem would be a fully automated classification system. Each image would be scanned and the proper classification code assigned to that image. In the mid 1980s, some research was begun in an effort to develop such an automated system. However, no viable approaches were ever fully developed.

In the past year, NIST has devoted resources to the problem of automating the classification process. By use of massively parallel neural network processing techniques, successful results have been obtained toward successfully computerizing the gross pattern classification recognition of fingerprint images. To date, the best accuracy achieved on a sample with equal numbers of five pattern types has been about 88% with 10% being unrecognizable. It is anticipated that a recognition system can be developed to achieve a 99% accuracy.

The speed to classify each image in this study, using massively parallel computing techniques, was .5 seconds/fingerprint. The architecture used for this study was an Active Memory Technology 510 Distributed Array Processor.⁵ With existing parallel computers, speeds sufficient to process the FBI's current workload can be obtained with a small number of systems. It should be noted that this speed is also essential for large scale testing of potential recognition methods. Extensive training, testing, and evaluation of potential recognition techniques will only be practical for researchers who have access to computer equipment with speeds capable of fulfilling the FBI current requirements. For example, operating at .5 seconds/fingerprint, about 7 hours would be required to automatically classify 50,000 fingerprint images (5000 ten-print cards). If parallel computing techniques were not available, an automatic classifier operating at 1000 seconds/fingerprint would require 578 days (operating 24 hours/day) to process the same 50,000 fingerprint images.

Work has also been initiated aimed at identifying core and delta locations. Once this has been accomplished, the next logical step will be to calculate the distance between the identified core and delta locations. This distance measure, in combination with the pattern type, will effectively divide the most common pattern types into several additional bins in the classification system.

However, the FBI should not limit itself to supporting research for this type of classification approach alone. As suggested by the C/M report, the FBI should establish a research program to investigate or develop alternative methods for fingerprint classification. New and successful approaches to the classification task may prove to be more cost-effective than pattern level classification. Any improvements to the classification process will reduce the number of matches required for each search request and therefore improve the overall performance of the IAFIS.

Approaches aimed at enhancing or replacing the conventional fingerprint pattern classification system may be pursued. Researchers would be free to use new algorithms for classifying images in any way and to any degree of fineness that proves effective. Their development would be unhampered by the current system. For example, schemes to calculate fingerprint classification codes, which may be unrecognizable to the human eye, could provide an accurate and

functionally viable classification technique when implemented on a computer.

8.4 CLASSIFICATION RECOMMENDATION

For a short-term solution, the usefulness of either or both of the first two approaches could be determined with a minimum amount of resources being expended and in a relatively short time frame. Any technique requiring a minimum of classifier effort to increase the number of classification bins and decrease the number in each bin would prove to be greatly beneficial to the IAFIS project in terms of cost and efficiency. These techniques (if successful) are designed to be implemented on a temporary basis. For the long-term perspective, research should be encouraged. Better or more efficient classification techniques developed and proven superior to existing methods can replace any interim system installed.

9. MINUTIAE DETECTION ISSUES

The image processing associated with minutiae identification, detection, and extraction is another very important area to consider for research. For the foreseeable future, it appears as if the FBI will continue to use a minutiae-based approach to fingerprint matching. Any matching algorithm chosen will continue to be dependent on the performance of the minutiae detection hardware and software.

Current equipment in use at the FBI detects between 30 to 35% of the true minutiae on a consistent and reliable basis. Although current commercial systems will detect up to 70% of the true minutiae, 30% are still not consistently detected. In the mid 1980s, NIST was working on developing matchers for latent use. The research came to a halt when it was discovered that the image minutiae available for testing were not reliable. Had the minutiae sets been of improved quality, improvements to latent matching might have been accomplished.

Since the detected minutiae are the input to all matching and verification processes, research in this area should receive special attention. Advances made in this area of minutiae detection will necessarily result in a better hit rate performance of any matcher that is used. As a consequence, the probability of a verifiable identification on a candidate drop from the matcher will increase. Thus, the emphasis for automated verification will be reduced.

10. RESEARCH DATABASES

As research efforts for each of the main fingerprint identification tasks are begun, standard sources of images are required to test these new approaches. Fingerprint image classification is one such area of research, where a large database of images is required for testing purposes.

For the open systems approach suggested by the C/M report to function properly, large databases of images must be available to researchers for developing and testing new techniques

or algorithms, and to the FBI for evaluating proposed approaches. The researcher must have access to a database for developing and training an algorithm. A second database is needed by the researcher to test the developed algorithm. An additional database shall be used by the FBI to evaluate the researcher's developed algorithm. For each of the major IAFIS processing tasks, these three tailored databases are required.

It should also be noted that if algorithm competition occurs, as suggested by the C/M report, a new and different test database shall be required for each competition.

Either the FBI or a contracted party should begin development of these sources of test data. One such database has already been completed. But additional development is still needed. Recently, a NIST-developed database designed to be used in image classification studies became available for acquisition by researchers, vendors, etc. The database contains two thousand pairs of 8-bit grayscale fingerprint images that were scanned, processed, and recorded on CD media.

Although this database fulfills an immediate need for sample fingerprint images, there are too few examples to adequately evaluate the high accuracy specifications required for IAFIS tasks. For example, a 99.7% accuracy rate is targeted for the automatic classification task. This implies that for every thousand images, a maximum of three images can be misclassified. To statistically test this requirement, enough samples must be made available to identify at least 100 erroneous classifications. Therefore, the database requires a minimum of 33 thousand images to adequately test an algorithm to be used for classification.

11. OTHER IAFIS TASKS

The following subsections will address some other IAFIS tasks, not specifically related to fingerprint identification operations, which deserve attention.

11.1 TEXTUAL DATA

The master plan calls for the Integrated Transmission Network (ITN) portion of the system to handle the electronic search requests. Part of each ITN transmission will contain several FBI required data fields for the subject's physical descriptors and demographic data. Input of textual data found on the front or back sides of the fingerprint card is a labor intensive task. Provision must be made for handling ten-print cards arriving through the mail. A reader station will be located in the mail room area to perform the conversion from card image to electronic image. Part of this conversion process will be to enter all textual fields into the computer. This input will range from well-formed typewritten characters to various qualities of handwriting.

For many of the submitting sites, this textual data may not have been stored on a computer. Rather than first entering the textual data into the computer, many of these sites may decide that they can accept a delay of a few days and mail the cards. This poses a potential bottleneck to

the FBI if the volume of mailed cards far exceeds the anticipated volume.

To respond to this potential problem, the mail room will need to have an adequate number of systems to process the incoming mailed cards. Furthermore, these systems should have the ability to perform OCR on the textual data fields of these cards. The OCR should be capable of handling not only typed fonts, but also hand printed and, perhaps, cursive data. If not already initiated, efforts should be started to identify existing OCR technology that could be utilized for this application.

11.2 IMAGE QUALITY

Currently every ten-print card that is sent to the FBI is inspected by the classifiers. When the ITN is implemented, this may cease to be the case if the gross pattern classification is transmitted along with each image. Although an image may be classifiable, its quality may be very poor. For this reason, an ongoing program of image quality control should be part of the IAFIS program. This program could take the form of a manual inspection system, a computer driven inspection, or a combination of the two.

Certain physical aspects of image quality are not very difficult to measure. Subjective quality measures pose a different set of problems. Although a person can look at an image and determine its quality, currently, there are no universally agreed upon metrics to judge fingerprint image quality automatically. For this reason, the FBI should give consideration to initiating serious research aimed at determining measures to automatically judge the quality of incoming transmissions.

11.3 GRayscale Requirements

Traditionally, it was believed that grayscale pixels should always be quantized to eight bits. In work recently performed at NIST, the eight bit quantization requirement has been questioned. For two randomly chosen fingerprints, hard copy outputs of each of the eight bit planes were produced. After examining each of these outputs, it appeared as if only four or five of the bit planes contained useful information. If only four bits of information were required per pixel, optical storage savings and other benefits could be substantial.

By no means was this a definitive experiment. However, this issue is probably worth pursuing. If this theory proved to be true, the amount of storage needed would be reduced. A large amount of additional study, with the cooperation of latent examiners and compression experts, would be required before this theory could be proved or disproved.

11.4 BACKUP REQUIREMENTS

Image capture of ten-print fingerprint images will begin before the IAFIS system is implemented. Data collected during this period must be retained for subsequent processing as improvements to the algorithms are developed. Due to cost, speed, and efficiency considerations, magnetic

tape media may be the best choice for this relatively short-term backup function and, perhaps, even for long-term backup purposes.

A decade ago, the basic tape backup system consisted of reel-to-reel, 9-track, 6250 CPI mylar-based tapes. Since that time, new magnetic tape technologies have been introduced which provide greater storage capacity while requiring less storage space. As the amount of data collected will be substantial and its integrity must be maintained, careful consideration should be given to the particular media chosen and the associated risks.

11.5 OPTICAL STORAGE

One of the most important components of the IAFIS system will be the on-line use and long-term storage of large numbers of fingerprint images used for verification purposes. These images are required for the final step of the identification operation. Without them, examiners will have no images to compare with the search request candidate drops. For this reason, the FBI should be directing attention to various aspects of optical storage systems.

Optical media and related technologies are evolving rapidly. Keeping abreast of this evolution is an essential aspect of a long term commitment to these technologies. Before any selection of an optical disk subsystem is made, the FBI should understand the available types of subsystems and media. Emerging standards in this field will assist in the acquisition of this information. Since the data to be stored on these optical disk subsystems are intended for long term use, the media chosen should be the best-fit for the required application. Therefore, the long term capabilities or life expectancies of the media must be one of the prime factors considered in the selection process.

The FBI should also be familiar with the proper care and handling procedures applicable to this media. Such techniques will minimize the potential occurrence of environmental, chemical, and mechanical failure mechanisms. Consideration should also be given to the operating environment where these systems will be located. After all of the factors associated with the optical storage application have been examined and trade-offs considered, the best choice for the FBI can be determined.

An appropriate aspect of optical disk subsystems to be considered relates to the monitoring of the on-line data. Information regarding media error rate and error distribution statistics will allow the FBI to monitor the status of their data. Approaches can be developed to perform off-line or real time on-line monitoring. From this information, proper recopying policies can be implemented to ensure continual data integrity.

11.6 LOCAL AND WIDE AREA COMMUNICATION STANDARDS

Localized clusters of computers are likely to exchange data via a local area network (LAN). The remote exchange between clusters (e.g., between a local agency and the state, or between a state and the FBI) is done via a wide area network (WAN) interface such as the X.25 protocol.

Rather than using proprietary LANs and private lines as a WAN, the FBI needs to consider emerging off-the-shelf products using standards, especially the Fiber Distributed Data Interface (FDDI) and Integrated Service Digital Network (ISDN). These specific standards are well suited for image exchanges. Research is needed for their specific use within the IAFIS architecture.

11.7 SOFTWARE STANDARDS

There are existing and emerging software standards for operating systems, databases and graphics that can invoke the above communication services. Standards such as GOSIP, POSIX, SQL and GKS need to be explored for permitting multi-vendor interoperability between various information systems. Security primitives and mechanisms together with a risk analysis are also necessary to protect against infiltration of an IAFIS node.

12. SUMMARY & CONCLUSIONS

On the whole, the Crawford and Mjolsness report presented an excellent overview of the current situation regarding the FBI's IAFIS project. The critical issues they presented are very realistic. However, their estimate of the consequences of failure may be on the optimistic side. Most of their recommendations should be supported by the FBI, especially those relating to keeping an open system approach and pursuing new research methods to solve problems of classification, image processing, and matching.

Although they recommend that the FBI pursue the conventional AFIS approach using a gross pattern classification, this would be a costly and inefficient approach. The volume of daily search requests and the size of their database makes a gross classification approach prohibitive. Before this decision is finalized, some short term research should be done to find a simple method of improving the classification process which would require very little additional classifier coding. The semi-automatic ridge counting and the core-to-delta distance measurements are two such approaches that have been described in this paper.

Exception must also be taken to their recommendation to conduct a yearly competition for the best algorithms. This could prove to be very cumbersome and wasteful of resources required to perform evaluations and implementations. A better approach is to update or replace existing algorithms when improved ones become available.

REFERENCES

- [1] Carol Crawford and Eric Mjolsness (October 1991). "Automated Fingerprint Identification: An Independent Study". This report provided to the FBI contains the details of their study.
- [2] The MITRE Corporation, "Systems Requirements Definition for the Integrated Automated Fingerprint Identification System", MTR-91W00069, Draft of May 1991.
- [3] The Workshop on the Electronic Exchange of Fingerprint Images, held in Gaithersburg MD, March 4-6, 1992 and sponsored by NIST was conducted to develop language to be included in the American National Standards Institute draft standard for the Interchange of Fingerprint information.
- [4] V. Vapnik, (1992). 'Principles of Risk Minimization for Learning Theory', in Advances in Neural Information Processing Systems 4, ed. John Moody et al., 831-838, Morgan Kaufmann.
David J.C. MacKay, (1992). 'Bayesian Model Comparison and Backprop Nets', in Advances in Neural Information Processing Systems 4, ed. John Moody et al., 839-846, Morgan Kaufmann.
I. Guyon, V.N. Vapnik, B.E. Boser, L.Y. Bottou and S.A. Solla (1992). 'Structural Risk Minimization for Character Recognition', in Advances in Neural Information Processing Systems 4, ed. John Moody et al., 471-479, Morgan Kaufmann.
John E. Moody, (1992). 'The Effective Number of Parameters: An Analysis of Generalization and Regularization in Nonlinear Learning Systems', in Advances in Neural Information Processing Systems 4, ed. John Moody et al., 847-854, Morgan Kaufmann.
- [5] DAP510c or equivalent may be identified in order to adequately specify the subject matter. In no case does such identification imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the equipment identified is necessarily the best available for the purpose.

BIBLIOGRAPHIC DATA SHEET

1. PUBLICATION OR REPORT NUMBER NISTIR 4892
2. PERFORMING ORGANIZATION REPORT NUMBER
3. PUBLICATION DATE AUGUST 1992

4. TITLE AND SUBTITLE

RESEARCH CONSIDERATIONS REGARDING FBI-IAFIS TASKS & REQUIREMENTS

5. AUTHOR(S)

R. McCabe, C. Wilson, D. Grubb

6. PERFORMING ORGANIZATION (IF JOINT OR OTHER THAN NIST, SEE INSTRUCTIONS)

U.S. DEPARTMENT OF COMMERCE
NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
GAITHERSBURG, MD 20899

7. CONTRACT/GRANT NUMBER

FBI Contract #AZ11 00-FBI-0001

8. TYPE OF REPORT AND PERIOD COVERED

FINAL

9. SPONSORING ORGANIZATION NAME AND COMPLETE ADDRESS (STREET, CITY, STATE, ZIP)

U.S. Department of Justice
Federal Bureau of Investigation
Room 1B268, Supply Technician
10th & Pennsylvania Avenue, N.W.
Washington, DC 20535

10. SUPPLEMENTARY NOTES

11. ABSTRACT (A 200-WORD OR LESS FACTUAL SUMMARY OF MOST SIGNIFICANT INFORMATION. IF DOCUMENT INCLUDES A SIGNIFICANT BIBLIOGRAPHY OR LITERATURE SURVEY, MENTION IT HERE)

The FBI is engaged in a revitalization effort of its Identification Division which will include a state-of-the-art Integrated Automated Fingerprint Identification System (IAFIS). To provide its users with the best possible IAFIS system, the FBI has solicited input from various sources. This report is primarily intended to provide a review of the programmatic needs and the directions of technical efforts that the Identification Division should consider. Alternative methods for solving current workload problems are presented. Candidate areas for IAFIS research are examined. Both long and short term suggestions are offered.

12. KEY WORDS (6 TO 12 ENTRIES; ALPHABETICAL ORDER; CAPITALIZE ONLY PROPER NAMES; AND SEPARATE KEY WORDS BY SEMICOLONS)

classification; fingerprint; grayscale; IAFIS; image quality; image processing; matching; minutiae; neural nets; relaxation algorithms; scale-space algorithms;

13. AVAILABILITY

UNLIMITED

FOR OFFICIAL DISTRIBUTION. DO NOT RELEASE TO NATIONAL TECHNICAL INFORMATION SERVICE (NTIS).

ORDER FROM SUPERINTENDENT OF DOCUMENTS, U.S. GOVERNMENT PRINTING OFFICE,
WASHINGTON, DC 20402.

ORDER FROM NATIONAL TECHNICAL INFORMATION SERVICE (NTIS), SPRINGFIELD, VA 22161.

14. NUMBER OF PRINTED PAGES

22

15. PRICE

A02

